

# 解决难以接近人群的抽样困境——应答推动抽样<sup>1</sup>

邱培媛, 杨洋, 周欢, 刘巧兰, 马骁

四川大学华西公共卫生学院健康教育, 四川成都 (610041)

E-mail: [antiaids@163.com](mailto:antiaids@163.com)

**摘 要:** 本文介绍了一种在难以接近人群中抽样的新方法——应答推动抽样 (Respondent-Driven Sampling, RDS) 的基本原理、抽样过程、与传统抽样方法相比所具有的优势及应用范围。鉴于应答推动抽样 (RDS) 方法在难以接近人群抽样中的优势及有效性, 此法将在诸如吸毒者、同性恋、商业性工作者, 以及流动人口等难以接近人群的调查研究中发挥重要作用。

**关键词:** 应答推动抽样; Respondent-Driven Sampling (RDS)

## 1. 引言

在调查研究中, 研究者通常会采用抽样调查的方法, 从总体中抽取一定数量的观察单位组成样本, 对样本进行调查, 并期望能用所得样本的信息推断总体。抽样包括概率抽样和非概率抽样。在概率抽样中, 由于观察单位被抽中的概率是已知的或是可以计算的, 因此可以计算抽样误差, 并对总体进行统计推论, 是研究者首选的抽样方法。常用的概率抽样包括单纯随机抽样、系统抽样、整群抽样和分层抽样等。概率抽样的前提是目标总体明确, 以及抽样框架明确。但是在研究中, 常常有些目标总体的抽样框架很难确定, 如吸毒人群、HIV 感染者、商业性工作者、同性恋等。这些特殊的群体被称作难以接近的人群。针对这类人群的调查研究, 概率抽样难以胜任, 现有研究通常采用非概率抽样方法, 但是非概率抽样同样也存在自身的局限性。

非概率抽样方法最常用的包括滚雪球抽样法、目标人群抽样法、时间-地点抽样法, 以及关键信息人抽样法。这些方法的优点在于无需事先构建抽样框架, 可用于难以确定目标总体的群体。但是, 非概率抽样所得样本的代表性和外推性却往往受到质疑。例如, 在滚雪球抽样中, 种子的选择对最终获得的样本的代表性有很大的影响, 如果种子的选取不是随机的, 那么最终获得的样本就有可能出现高同质性, 或是某一特征人群的过度抽样<sup>[1]</sup>。尽管非概率抽样在难以接近的人群中得到广泛的应用, 方法也发展得较为成熟, 但是由于非概率抽样中观察单位被抽中的概率是未知的, 不能计算抽样误差, 因此不能通过样本信息推断总体, 其结论的外推也就受到限制。

## 2. 应答推动抽样 (Respondent-Driven Sampling, RDS)

针对难以接近人群研究中面临的抽样困境, Douglas D. Heckathorn 在 1997 年发表的一篇论文中提出了一种新的抽样方法——应答推动抽样 (Respondent-Driven Sampling, RDS)<sup>[2]</sup>。RDS 方法是在社会网络分析方法发展的基础上将经典的滚雪球等链式推举法 (Chain-Referral Sampling) 改良而形成的, 不仅具有概率抽样特性, 而且能借助社会网络分析方法对研究对象的社会网络特征进行分析。

<sup>1</sup>本课题得到高等学校博士学科点专项科研基金 (项目编号: 20060610087), 国家自然科学基金 (项目编号: 70673067) 的资助。

## 2.1 原理与方法

### 2.1.1 RDS方法和随机抽样法比较

随机抽样是按照随机原则，从总体中抽取一定数量的个体形成样本，再通过样本调查获取的信息推断总体（图1（a））。因为总体中每个个体被抽中的概率相等，因此样本对总体有较好的代表性。但是当总体的大小和边界不能确定时，如吸毒者，商业性性工作者，流动人口等难以接近人群的总体，随机抽样的应用就受到限制。

与常规随机抽样的原理不同，RDS并不通过样本信息直接推断总体，而是通过样本获得社会网络的信息，并应用社会网络分析方法得到抽样概率，计算抽样误差，利用样本形成的社会网络的信息推断所研究的总体<sup>[3]</sup>（图1（b））。RDS方法不需要预先设定抽样框架，因此可应用于上述难以接近的人群。

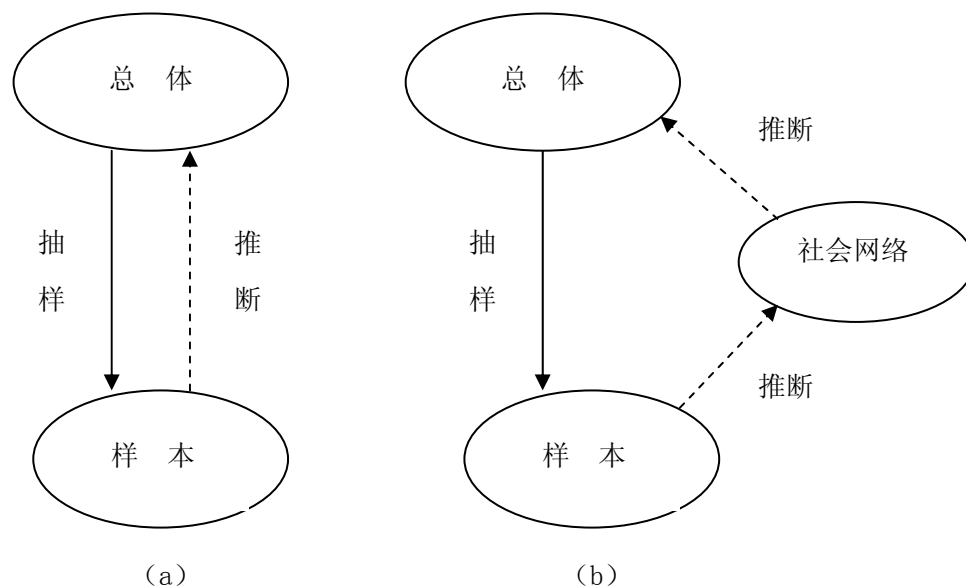


图1 常规方法与RDS方法的抽样和统计推断比较

(a) 为常规方法的抽样和统计推断；(b) 为RDS方法的抽样和统计推断

### 2.1.2 RDS方法和链式推举法

和经典链式推举法类似，RDS方法也是首先在目标人群中选取一定数量的种子，然后通过种子获得一级抽样人群，再通过一级抽样人群获得二级抽样人群，在 $n$ 级抽样后，最终会得到一个稳定的目标人群的样本。推选同伴的数量一般限制在2-4人/人。应用马尔可夫链的理论<sup>[4,5]</sup>和社会网络分析方法<sup>[6-9]</sup>可以证明，当 $n=4\sim 6$ 时，抽样获取的样本特征就会趋向稳定并且对所研究的总体有较好的代表性。

## 2.2 RDS方法的优势和应用范围

### 2.2.1 RDS方法的优势

#### 1、可计算抽样概率，对总体进行推断

RDS过程中调查了每个研究对象的社会网络特征，应用社会网络的方法可以计算抽样概率，估计抽样误差，对总体进行推断。并可通过社会网络的方法，对调查对象的社会网络进行调查，为了解调查对象提供更多的信息。

## 2、在RDS方法中，种子是否为随机抽样获得对最后得到的样本的代表性没有影响

马尔可夫链理论可以证明<sup>[5]</sup>，通过RDS方法进行抽样，在样本达到平衡后，样本的代表性将不受最初选择种子的影响。同时应用马尔可夫链理论也可以证明，样本达到平衡的速度很快，通过4~6轮抽样，样本即可达到平衡。

## 3、RDS方法能避免过度抽样

RDS方法限定了推选同伴的数量，限定每个人只能推选一定数量的同伴（一般为2~4人），使得推举链延长，避免了某一特征人群的过度抽样<sup>[10]</sup>。

## 4、RDS包含了双重激励机制

RDS方法的激励机制分为两步，第一步是对研究对象参与调查的奖励，第二步是对研究对象推荐同伴的奖励。有研究表明，双重激励有两方面的作用，一方面是通过物质奖励，另一方面是让那些对物质奖励不感兴趣的个体屈服于同伴压力，从而提高样本的收集效率<sup>[5]</sup>。

### 2.2.2 RDS 方法的应用范围

尽管 RDS 方法在难以接近人群的抽样中具有优势，但并不是适用于所有的难以接近人群。RDS 方法只能应用于社会网络已存在的群体中，如吸毒者、流动人口等，而不能应用于社会网络不存在群体中，如自杀群体。

## 3. 小结

RDS 方法无需事先构建抽样框架，同时又具有概率抽样的特性，可计算抽样概率，推断总体。因此，RDS 方法在难以接近人群中的应用既优于传统的随机抽样方法，也优于非概率抽样方法。以康奈尔大学等为代表的国外研究单位近来将 RDS 方法应用于吸毒人群的调查<sup>[6]</sup>、HIV 监测和咨询<sup>[7]</sup>、商业性工作者研究<sup>[8]</sup>、以及爵士乐手研究等，已经表明其有效性。但是在国内还鲜有应用 RDS 方法的研究报告。鉴于应答推动抽样（RDS）方法在难以接近人群抽样中的优势，此法将在我国有关吸毒者、流动人口等调查研究中发挥重要作用。

## 参考文献

- [1] Douglas D. Heckathorn. Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Population [J]. Social Problems, 2002, 49(1):11-34
- [2] Douglas D. Heckathorn. Respondent-Driven Sampling: A New Approach to the Study of Hidden Population [J]. Social Problems, 1997, 44(2):174-198
- [3] Matthew J. Salganik, Douglas D. Heckathorn. Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling [J]. Sociological Methodology, 2004, 34: 193-239
- [4] 袁方主编. 社会研究方法教程[M]. 北京: 北京大学出版社, 1997: 653-655
- [5] Douglas D. Heckathorn, Salaam Semaan, Robert S. Broadhead et al. Extensions of Respondent-Driven Sampling: A New Approach to the Study of Injection Drug Users Aged 18-25 [J]. AIDS and Behavior, 2002, 6:55-67
- [6] Briadhead, Robert S., Douglas D. Heckathorn, Jean-Paul C. Grund. Drug users versus outreach workers in combating AIDS: Preliminary results of a peer-driven intervention [J]. Journal of Drug Issues, 1995, 25(3):531-564
- [7] Briadhead, Robert S., Douglas D. Heckathorn. AIDS prevention outreach among injection drug users: Agency problems and new approaches [J]. Social Problems, 1994, 41:473-495
- [8] Mikael Jansson, Cecilia Benoit. Some Benefits and Challenges in Conducting Community-Academic Research on Youth Involved in the Sex Trade [J]. Presented at the conference: Making the Connection: Sexually Exploited Youth, the Challenges and Opportunities. May 28-30, 2002, Victoria, BC, Canada
- [9] Salaam Semaan, Jennifer Lauby, Jon Liebman. Street an Network Sampling in Evaluation Studies of HIV Risk-Reduction Interventions [J]. AIDS Review, 2002(4):213-223
- [10] Robert Magnani, Keith Sabin, Tobi Saidel et al. Review of Sampling Hard-To-Reach and Hidden Population for HIV Surveillance [J]. AIDS, 2005, 19(suppl 2):S67-S72

## **A Sampling Method Using in Hard-to-Reach Population — Respondent-Driven Sampling**

Qiu Peiyuan, Yang Yang, Zhou Huan, Liu Qiaolan, Ma Xiao

Department of Health Education, School of Public Health, Sichuan University, Chengdu, Sichuan  
(610041)

### **Abstract**

This paper introduced a new sampling method, respondent-driven sampling (RDS), which is used in hard-to-reach population, including its principle, sampling process and advantages comparing to commonly used sampling methods. Taken these advantages into consideration, RDS method will play an important role in obtaining representative sample in hard-to-reach population, such as Injection Drug Users, Commercial Sexual Workers, Homosexual Population, Migrant Population and so on.

**Keywords:** Respondent-Driven Sampling (RDS)